



Reenacting sensorimotor features of drawing movements from friction sounds

Etienne Thoret, Mitsuko Aramaki, Richard Kronland-Martinet, Jean-Luc
Velay, Sølvi Ystad

► To cite this version:

Etienne Thoret, Mitsuko Aramaki, Richard Kronland-Martinet, Jean-Luc Velay, Sølvi Ystad. Reen-acting sensorimotor features of drawing movements from friction sounds. From Sounds to Music and Emotions, Springer-Verlag Berlin Heidelberg, pp.130-153, 2013, Lecture Notes in Computer Science, 978-3-642-41248-6. 10.1007/978-3-642-41248-6_8 . hal-00866358

HAL Id: hal-00866358

<https://hal.science/hal-00866358>

Submitted on 26 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reenacting Sensorimotor Features of Drawing Movements from Friction Sounds

Etienne Thoret¹, Mitsuko Aramaki¹, Richard Kronland-Martinet¹, Jean-Luc Velay², and Sølvi Ystad¹

¹ LMA - CNRS - UPR 7051 – Aix Marseille Univ. – École Centrale Marseille

² LNC - CNRS - UMR 7291 – Aix Marseille Univ.

{thoret,aramaki,kronland,ystad}@lma.cnrs-mrs.fr
jean-luc.velay@univ-amu.fr

Abstract. Even though we generally don't pay attention to the friction sounds produced when we are writing or drawing, these sounds are recordable, and can even evoke the underlying gesture. In this paper, auditory perception of such sounds, and the internal representations they evoke when we listen to them, is considered from the sensorimotor learning point of view. The use of synthesis processes of friction sounds makes it possible to investigate the perceptual influence of each gestures parameter separately. Here, the influence of the velocity profile on the mental representation of the gesture induced by a friction sound was investigated through 3 experiments. The results reveal the perceptual relevance of this parameter, and particularly a specific morphology corresponding to biological movements, the so-called 1/3-power law. The experiments are discussed according to the sensorimotor theory and the invariant taxonomy of the ecological approach.

Keywords: Sensorimotor Approach of Auditory Perception - Friction Sounds - 1/3 power law - Biological Movement

1 Introduction

The relation between sound and movement is a very wide field of research. In this article we will focus on a particular topic related to this field namely the relation between a sound and a specific movement: the human drawing movement. Evoking a movement with a monophonic source only by acting on timbre variations of the sound is a process often used by electroacoustic musicians and sound engineers. Musicology analyses proposed semiotic descriptions of perceived movements in musical pieces [10]. Moreover, the general relations between intrinsic sound properties and movements have been tackled in previous studies by Adrien Merer in [27] and [28]. The motions evoked by monophonic sounds were investigated from two angles, first by using a free categorization task, with so-called abstract sounds, that is, sounds which source was not easily identifiable. And secondly with a perceptual characterization of these evoked motions by studying the drawings produced by a group of subjects using a purpose graphical user interface. These studies had a very interesting approach which enabled

to extract relevant perceptual features of relations between timbre variations of a sound and the evoked movements, and led to a sound synthesizer of evoked movements based on semantic and graphics controls.

Different psychological studies tackled the problem of sound event recognition. They principally based on the ecological approach of visual perception introduced by Gibson [15] which supports the idea that perception emerges from the extraction of invariant features in a sensory flow, and moreover from the organization of the perceptual system itself. This approach has been formalized for the auditory perception in different studies [46, 12, 13]. Opposed to this view, the information theory proposed that the perception is the result of a process with multiple steps which enables the association between a memorized abstract representation, and its identity and signification. In [29], McAdams has an intermediate position, he proposed to adopt the point of view of the information theory, and the notion of auditory representation, but keeping the terminology of invariants features which comes from ecological approach. It is well adapted to the description of the material world, and particularly in highlighting that some properties are perceived as invariant when others can change without changing the perception and signification of the stimulus. Moreover, we can argue that as it essentially concerns the recognition of sound events, it is adapted to adopt a representationalist view with this terminology to describe the information which is used to compare a representation of a stimulus with memorized representations. It is therefore proposed that the acoustic properties that carry information that enables the recognition of a sound event can be defined as structural and transformational invariants. The information that enables to identify the nature of the sound source was defined as a structural invariant. For instance, it has been shown that impact sounds contain sufficient information to enable the discrimination between the materials of impacted objects [47, 18, 2]. The information that specifies the type of change is known as a transformational invariant. For instance, a study revealed that the rhythm of a series of impacts enables to predict if a glass will break or bounce [46].

In the following study, we will focus on a particular type of sound event, the sound produced by the friction between a pen and a paper when we are drawing. This sound is audible but we did not necessary pay attention to it. The timbre variations contained in it may enable to imagine a movement and therefore a gesture. Are we able to recognize the gesture from the friction sound? If yes, can we imagine the shape which has been drawn? A graphical gesture can be mainly defined as a couple of velocity and pressure profiles which induce changes in the produced friction sound. In the following we will focus on the velocity profile of the gestures. We will investigate whether this information is sufficient to recognize a gesture and if it can be considered as a transformational invariant concerning human graphical gestures.

The relation between sound and gesture can be approached regarding a general theory at the edge between philosophy and cognitive sciences, called *en-action*, which was introduced by Francisco J. Varela [37, 38]. This theory proposed a new approach of cognition distinct from the classical top down and

bottom up models coming from cognitivist and connectionist approaches. He reintroduced actions and intentions at the center of our conception of cognitive processes involved in the perception of a stimulus. Varela defined the notion of *incarnated/embodied actions* which can be summed up by the main idea that our perceptual processes are modeled by our actions and intentions, and that actions are central and cannot be separated from perception. Regarding the invariant taxonomy, it can be hypothesized that invariants which are used by the perceptual processes to identify a sound event refer to embodied properties of these actions [37]. As mentioned before, it should still be noted that invariant taxonomy comes from the ecological approach which is not consistent in many points with the enactive theory, but in this study, we will consider the notion of invariant as the information used by perceptual processes to identify and to recognize an event. The low level coding of embodied action has been supported by functional imagery observations in monkeys which revealed the existence of specific neurons in the ventral premotor cortex, the *so-called* mirror neurons, which fired either when the monkeys make an action or when they just observe it [32, 11]. These observations have also been done in monkeys in the case of the auditory modality [19]. Finally, other electrophysiological and anatomical observations have been done with musicians whose brain area involved in the practice of their instrument was activated when they just listened to the instrument. Moreover, it has been shown that the intensity of activation is higher according to the musician's degree of expertise [3]. These last observations highlighted the importance of the perception-action coupling, also called the sensorimotor loop, in the perceptual processes and particularly in the case of auditory perception.

In this paper, we will investigate the relation between a friction sound produced by someones drawing and the evoked movement with the previous embodied action approach. It enables to make strong hypothesis about the dynamic features which can be imagined from a friction sound. We will aim at highlighting which parameters of the gesture can be evoked through a friction sound. Here we focus on the velocity profile, to set up experiments which investigate this question, we need friction sounds produced by a specific velocity profile. A graphic tablet and a microphone can be used for this purpose. This solution enables the analysis of the sound regarding the gesture properties but doesn't provide the possibility to control precisely the velocity of the writer. For control purposes, it would be more interesting to create synthetic friction sounds from given velocity profiles. A synthesis process of friction sounds which enables to synthesize such friction sounds will be present in a following section.

Finally, we will investigate the representation of a gesture from a friction sound in three experiments in which both recorded and synthesized friction sounds are used. In the first two, friction sounds produced when a writer draws different shapes will have to be associated to static visual shapes to identify if friction sounds can evoke a specific gesture, and furthermore a geometric shape. A third experiment investigates the relevance of a biological relation which links the velocity of a human gesture to the curvature of the trajectory from the audi-

tory point of view. The results of these experiments will be discussed according to the sensorimotor theoretical background finally.

2 A Synthesis Model of Friction Sounds

In the three experiments which will be presented, a part of the stimuli will be generated with a sound synthesis process. The main goal will be to evaluate the relevance of the velocity profile in the representation of a gesture underlying a friction sound: Are we able to imagine the gesture made only by listening the friction sound? Can we even recognized the shape which is drawn from the sound? And at last, is there morphologies, *so-called* invariants, linked to the velocity profile which enable the recognition of a human gesture from an auditory point of view?

From a synthesis point of view, a paradigm well adapted to the invariant's taxonomy is the *action/object* paradigm. It consists in defining the sound as the result of an action on an object (e.g. "rubbing on a metal plate"). A natural way to implement the action/object paradigm is to use subtractive synthesis, also called a source filter approach. This method enables to separate synthesis of the action, the exciter, *e.g. the transformational invariant*, and the object, the resonator, *e.g. the structural invariant*. To synthesize friction sounds with this approach, we used physically informed model, also called, phenomenological model presented by Gaver in [12] and improved by Van den Doel in [36]. It aims at reproducing the perceptual effect rather than the real physical behavior.

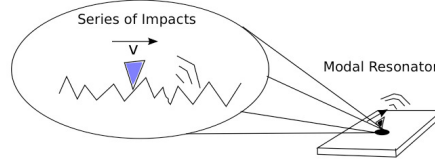
This approach considers a friction sound as the result of a series of impacts produced by the interaction between the pencil mine and the asperities of the surface, see Figure 1. With a source-resonator model, it is possible to synthesize friction sounds by reading a noise wavetable with a velocity linked to the velocity of the gesture and filtered by a resonant filter bank adjusted to model the characteristics of the object which is rubbed or scratched (see figure 1) [7, 8]. The noise wavetable represents the profile of the surface which is rubbed. Resonant filter bank simulates the resonances of the rubbed object and is characterized by a set of frequency and bandwidth values [1, 2].

This synthesis model is particularly well tuned for our study, it indeed enables to generate a synthetic friction sound which varies only according to the velocity of the gesture.

3 A Relevant Acoustical Information: the Timbre Variations due to the Velocity of the Pen

Graphical tablets henceforth allowed to accurately record dynamical information like velocity and pressure, and to use it for comparing two shapes according to the kinematics which have produced them. As evoked before, many studies have highlighted the importance of the velocity profile in the production of a movement, and in particular, of graphical movements. Moreover, the friction sound

(A) Phenomenological Model



(B) Source-Filter Approach

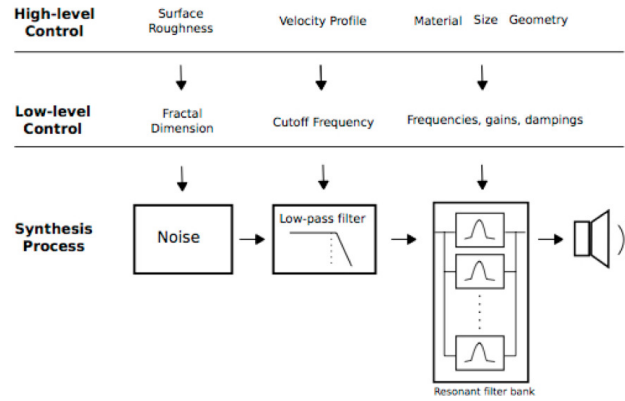


Fig. 1. Panel A: Physically Informed Friction Sound Synthesis Model - The friction sound is assumed as a series of impact of a plectrum, in our study the pen, on the asperities of a modal resonator – Panel B: Implementation of the phenomenological model with a source-filter approach that enables to separate the action, here the gesture, the object, here a paper on a table. The different levels of control are presented. The high level one corresponds to the intuitive control proposed to a user which enables to define an object from a semantical control of its perceived material and shape, while the low level corresponds to the synthesis parameters.

synthesis previously presented enables to synthesize the friction sounds produced when someone is drawing based on the velocity profile only. As mentioned in the introduction, the velocity profile is a very important characteristic of a gesture, which may be involved at different levels of perception of a biological movement both in the visual system [41, 43] and in the kinesthetic one [45]. Here we aim at investigating if this parameter is also a relevant cue to identify a drawn shape from a friction sound.

Recording Session. We asked someone to draw six different shapes on a paper, see Figure 3. While the velocity profile of the drawing movement was recorded

thanks to a graphic tablet³, the friction sound produced by the pen on the paper was also recorded with a microphone, see Figure 2 for the experimental set up. The writer was asked to draw each shape as fluidly and as naturally as possible. Empirical observations were made just by listening to the recorded friction sounds. The circle seems to be a very particular shape. It indeed has a very uniform friction sound, with little timbre variations, while the ellipse, arches, line and loops have more important ones. The lemniscate seems intermediate between the other shapes, it indeed has a sound which contains more variations than the circle, but less than the loops, the arches and the line. Among the shapes which have a lot of timbre variations like ellipse, loops, line and arches, it should be noted that the line and the arches are distinct from the loops and the ellipse actually. They contain cusps, which imply silences in the friction sounds which are very audible and provide important perceptual cues linked to the geometry of the drawn shape.

Dealing with these empirical considerations, we chose to establish two corpuses of four shapes, one composed of shapes that are *a priori* distinct, the ellipse, the circle, the line and the arches. The second one of shapes which are *a priori* closer, i.e. the ellipse, the loops, the lemniscate and the circle. In particular, the first one has shapes which contain cusps: line and arches. A period of the velocity profile for each shape is presented in Figure 4, for the arches and the loops, only one period of the four shapes is presented. The circle has a specific velocity profile with a velocity almost constant during the whole drawing process. Otherwise, the durations of one period vary considerably according to the different shapes. To avoid this problem, the stimuli which have been chosen in the following contained four periods for the ellipse, the lemniscate and the line (one could say four *round trips* for this one). For the circle, only two periods were chosen, since the global duration for one period was indeed longer than for the other shapes. The whole durations of the chosen stimuli are summarized in Table 1.

Table 1. Durations of the Performances Chosen for the Recorded Velocity Profiles and Friction Sounds (in Seconds)

| Circle | Ellipse | Arches | Line | Lemniscate | Loops |
|--------|---------|--------|------|------------|-------|
| 5.2 | 5.8 | 5.1 | 5.2 | 5.6 | 5.4 |

A way to formally compare shapes of the two corpuses according to a dynamical dimension is to compare the proximity of the recorded velocity profiles with a metric. Thus, in a second time, to set up listening tests with friction sounds as stimuli to establish a perceptual distance between pairs of shapes. At last, the classifications computed from these mathematical and perceptual distances can

³ Wacom Intuos 3

be compared to evaluate if the velocity profile provides a relevant information from an auditory point of view.



Fig. 2. Experimental recording set-up

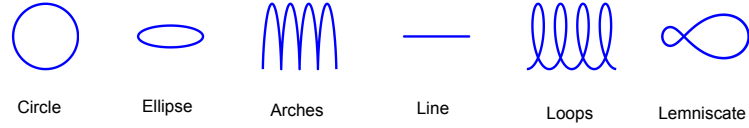


Fig. 3. The six shapes chosen for the experiments

3.1 Clustering of Geometrical Shapes from the Velocity Profile

Practically, recorded velocity profiles correspond to series of N measurements v_i at the sample frequency of the graphic tablet (here 200 Hz). A velocity profile is then defined as an array of N points according to the duration of the drawing. Finally, to compare two velocity profiles v and w it is necessary to be able to compare two vectors of different lengths. The durations of two drawings is indeed most of the time different for different shapes.

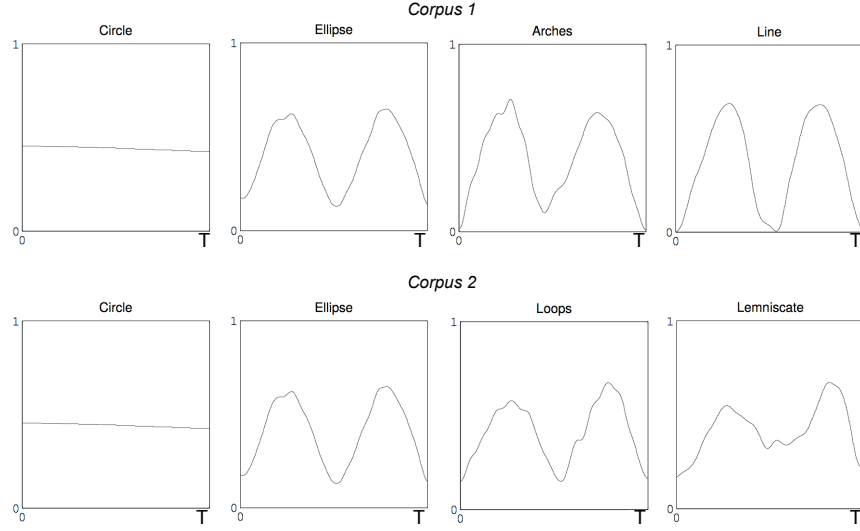


Fig. 4. Periods of the velocity profiles normalized and resampled on 512 samples used to compute the clustering

Euclidean Distance Between Two Velocity Profiles. A common mathematical tool used to compare two vectors is the inner scalar product that enables to define a distance according to a metric. The choice of the metric is crucial. It indeed defines the way the distance between shapes will be calculated, and it defines an objective measure between two shapes in terms of velocity profiles. The most classical metric is the euclidean one which corresponds to the following inner product between two vectors v_i and w_i , of the same length:

$$\langle v|w \rangle = \sum_{k=1}^N v_k w_k \quad (1)$$

The distance between two velocity profiles can then be obtained from the Euclidean distance, $d(v, w) = \|v - w\| = \sqrt{\langle v - w | v - w \rangle}$, which is minimal when $v = w$ and increases as the difference between v and w increases. In the case of velocity profiles, since arrays are of different lengths. It has been chosen to resample each velocity profile in 512 samples and to normalize them according to their mean value. The rationale is that the recordings are about the same duration, see Table 1. Thus, this normalization does not introduce a bias in the calculus of the distance. More complex algorithms exist to compute a distance between two arrays of different lengths, such as Dynamic Time Warping [20]. This last one is effective but very expensive in computing time and provide no significant advantages here from the resampling.

Dissimilarity Matrix and Clustering. The Euclidean distance enables to compute a distance between each pair of shape for each corpus (6 pairs for

each corpus). And moreover, to create a dissimilarity matrix D in which each cell represents the distance between the two velocity profiles associated to two shapes. The diagonal values of this matrix are equal to 0, indeed the distance between two equal velocity profiles is null. Two hierarchical clustering analysis of D , with complete linkage, were then effectuated from the dissimilarity matrices of each corpus. The dendrograms corresponding to this matrices are presented in Figure 5. A dendrogram corresponds to a hierarchical tree which represents a classification of different objects, here the velocity profiles. The height of each U-shape represents the distance between two objects or two clusters. For the two corpuses, the clusterings confirm the empirical observations, the circle is the shape that is most different from the others in the two sets of shapes. This could be explained by its velocity profile that is almost constant. In the first corpus, it is noticeable that the arches seem to be about equally distant from the ellipse and the line. In the second corpus, as expected, the ellipse and the loops are very close while the lemniscate is intermediate between the circle and the other two shapes. In order to determine if the previous classification obtained from the velocity profiles is relevant from an auditory perceptual point of view, two listening tests have been set up.

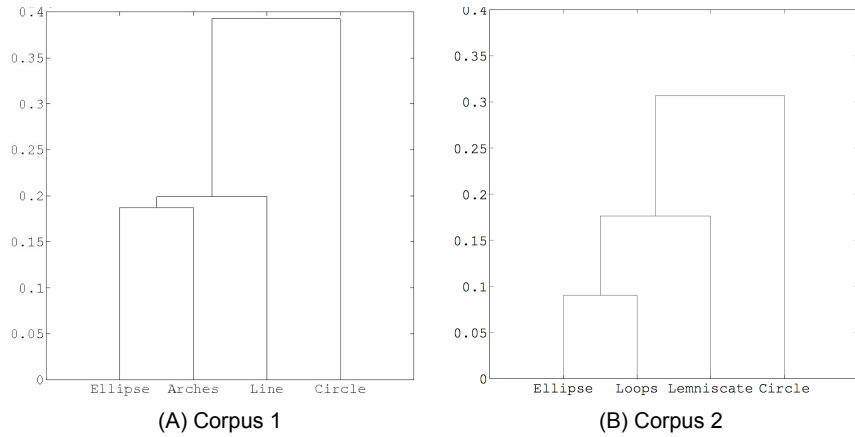


Fig. 5. Panel A and B: Ascending hierarchical clustering computed from the dissimilarity matrices computed from the corpus 1 and 2 respectively

3.2 Clustering of Geometrical Shapes from Perceptual Comparisons of Friction Sounds

The previous mathematical clustering based on the velocity of the gesture made to draw the shapes enables to evaluate the proximity between the shapes of

the two corpuses from an *objective* point of view. Our aim is to evaluate if the velocity profile is also a relevant information to compare two shapes *from the perception of friction sounds*. In other words, to investigate if the velocity profile conveys information about a gesture from the auditory point of view.

In the following, two listening tests with recorded and synthesized sounds produced during the drawing that were to be associated to the four shapes of each corpus are presented. Clusterings can then be obtained from the results of the listening tests and compared with the mathematical ones obtained previously. In order to establish a perceptual clustering of the shapes, the two listening tests consist in a association test where subjects have to associate the friction sounds to the correct shapes.

Experiment 1 – Distinct Shapes

Subjects. Twenty participants took part in the experiment: 9 women and 11 men. The average age was 30.65 years (SD=13.11). None of the subjects were familiar with the topic of the study before the test.

Stimuli. The first listening test deals with the shapes of the corpus 1 which contains the most distinct shapes with regard to the velocity profile. The auditory stimuli are composed of eight friction sounds, four recorded and four synthesized, obtained from the shapes of corpus 1 collected during the recording sessions presented previously. The synthesized sounds were generated with the friction sound synthesis model previously presented.

Task. The subjects were asked to univocally associate four friction sounds – among the four available – to the four shapes. The test was composed by eight trials: 2 types of sound x 4 repetitions.

Results and Short Discussion. For each subject and each type of sound – synthesized vs. recorded – sixteen scores of association between a sound and a shape were averaged across the four trials. All the results were stored in confusion matrices. The global averaged scores of success of the four shapes for recorded and synthesized stimuli are presented in Table 2. The results are clear, each friction sound has been properly associated to the corresponding shape above a random level⁴. Moreover the synthesized stimuli provide results which are not significantly different from the recorded ones which confirms the hypothesis about the perceptual relevance of the velocity profile.

Experiment 2 – Close Shapes

Subjects. Eighteen participants took part in the experiment, 8 women and 10 men. Their average age was 31.56 years (SD=13.73). None of the subjects were familiar with the topic of the study prior to the test.

⁴ The random level is defined at a 25% sound to shape association rate.

Table 2. Scores of success for recorded and synthesized stimuli of corpus 1 averaged across subjects – Mean and Standard Error in Percentages – Scores higher than statistical chance are bolded.

| | Circle | Ellipse | Arches | Line |
|-------------|----------------------|----------------------|---------------------|---------------------|
| Recorded | 98.75 1.25 | 81.25 6.25 | 80. 6.44 | 87.5 1.72 |
| Synthesized | 98.75 1.25 | 87.5 4.97 | 82.5 5.76 | 97.5 3.08 |

Stimuli. The second listening test deals with the shapes of corpus 2 which have shapes with closer geometries. Auditory stimuli are composed by eight friction sounds obtained from the shapes of corpus 2, four recorded and four synthesized, collected during the recording sessions presented previously. As for the experiment 1, the synthesized stimuli are generated with the friction sound synthesis model presented previously.

Task. The task was the same as in Experiment 1.

Results and Short Discussion. The data analysis is the same as in the previous experiment. The results reveal that, except for the loops, each sound was associated with the correct shape with a success rate above random level. Only the recorded loops were not recognized above chance. The scores of success are summarized in Table 3. Confusions appear between the ellipse and the loops, the score of association between these two shapes is not significantly different which means that they were confounded, see Table 3.

Clustering Analysis. The two previous listening tests revealed that when shapes are sufficiently different, it is possible to discriminate them simply from friction sounds. To valid entirely this statement, an additional analysis is necessary. A perceptual distance matrix between shapes was therefore computed from the confusion matrices obtained in the two experiments. They were firstly symmetrized, to implicitly merge the rate of association of sound i to shape j with the rate of association of sound j to shape i into one value representing the perceptual distance between the two shapes. The symmetrized confusion matrix \tilde{C} is obtained by:

$$\tilde{C} = \frac{C + C^t}{2} \quad (2)$$

with C^t the transposed version of matrix C . Then a discrimination matrix \tilde{D} was obtained by $\tilde{D} = 1 - \tilde{C}$. At last a pairwise distance matrix D is computed with the Euclidean metric.

For each corpus, two discrimination matrices are computed: one for the recorded sounds, and one for the synthesized one. Like for the mathematical

Table 3. Scores of success for recorded and synthesized stimuli of corpus 2 (A) — Scores of association between loops and ellipse for recorded and synthesized stimuli (B). All the scores are averaged across the subjects. Mean and standard error are presented in percentages. Scores higher than statistical chance are bolded.

| (A) | | | | |
|-------------|----------------------|----------------------|----------------------|----------------------|
| | Circle | Ellipse | Lemniscate | Loops |
| Recorded | 97.22 2.78 | 41.67 7.29 | 68.06 8.04 | 29.17 7.36 |
| Synthesized | 100. 0. | 50. 4.52 | 81.94 6.00 | 43.06 6.00 |

| (B) | | |
|-------------|---------------------------------------------|---------------------------------------------|
| | Ellipse Sound \leftrightarrow Loops Sound | Loops Sound \leftrightarrow Ellipse Sound |
| Recorded | 51.39 6.22 | 45.83 7.89 |
| Synthesized | 45.83 5.05 | 43.06 4.87 |

dissimilarity matrix, an ascending hierarchical clustering are computed from these discrimination matrices and provide dendrograms. Four clusterings are made from the whole results, see Figures 5 and 7. A global observation of the dendrograms lead us to hypothesize that, for the two corpuses, the two perceptual shapes classifications, synthesized vs. recorded, are equivalent. Indeed, in each case, the relative rank of proximity between whole shapes are the same. To statistically validate this, it is necessary to introduce the notion of cophenetic distances.

The problem of comparing two dendrograms has already been studied in phylogenetics. One goal of this field of biology is to understand the evolution of living beings according to molecular sequencing and morphological data which are collected into dissimilarity matrices and presented with dendrograms. Thus, the comparison of dendrograms has been tackled to compare morphological and molecular observations. As previously presented, a dendrogram is a representation of distances between different objects, and the composition of the clusters of the dendrogram is made according to a specific metric. A dendrogram is then characterized by distances between clusters, in which a specific distance has been defined and is called the cophenetic distance⁵ [35].

⁵ According to the help of the function *cophenet* in Matlab©, the cophenetic distance can be defined as:

An example of dendrogram and the associated cophenetic distances are presented in Figure 6. The cophenetic distances are sorted in an array for each dendrogram. To determine whether two dendrograms are statistically equivalent, it has been proposed to compute the Pearson's and Spearman's correlation coefficients between the two arrays of cophenetic distances. The Pearson's correlation coefficient r corresponds to a quantitative comparison of the linear correlation between shapes. And the Spearman's correlation coefficient ρ , corresponds to a qualitative comparison of the clusterings which takes into account of the ranks of the cophenetic distances between shapes.

We wanted to compare the two dendrograms obtained in each listening test. With the statistical method presented here, no significant differences are observed both for the experiment 1 and 2. The correlation coefficients are presented in Figure 7.

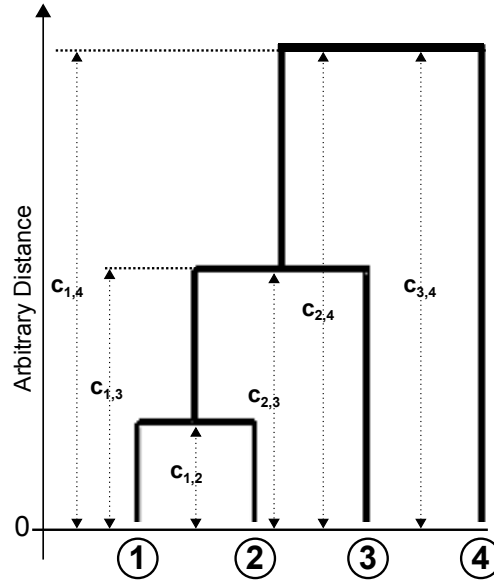


Fig. 6. An example of dendrogram and the associated cophenetic distances $c_{i,j}$ between the different objects – The cophenetic distances are summed up in an array: $C = [c_{1,2}, c_{1,3}, c_{1,4}, c_{2,3}, c_{2,4}, c_{3,4}]$ – and can be compared with cophenetic distances of others dendrograms with Pearson's and Spearman's correlation coefficients.

The [...] distance between two observations is represented in a dendrogram by the height of the link at which those two observations are first joined. That height is the distance between the two subclusters that are merged by that link.

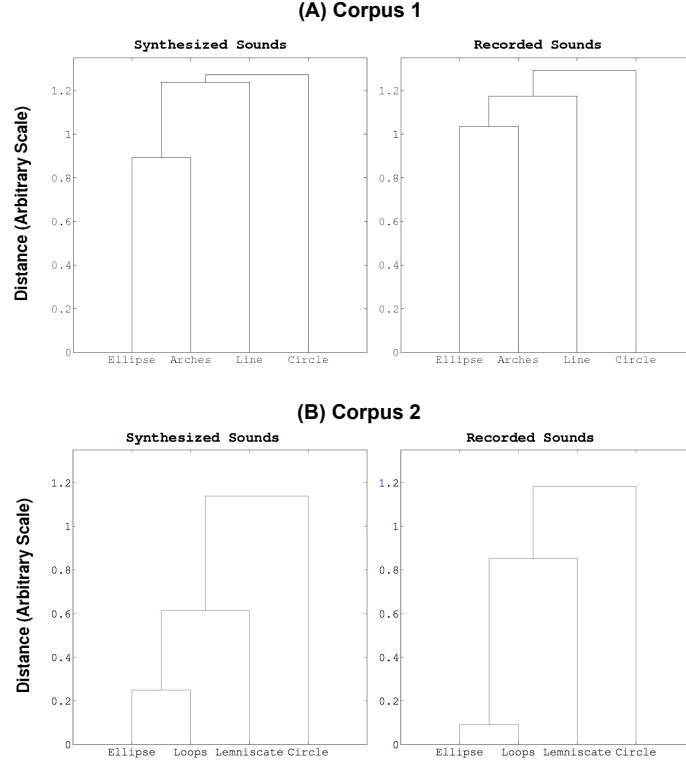


Fig. 7. Panel A: Ascending hierarchical clustering computed from the confusion matrices of the experiment 1 - Significant correlations were found between cophenetic distances of the two clusterings ($r = .89$ and $\rho = 1.$) — Panel B: Ascending hierarchical clustering computed from the confusion matrices of the experiment 2 - Significant correlations were found between cophenetic distances of the two clusterings ($r = .94$ and $\rho = 1.$) — All correlation coefficients are significant.

3.3 Comparison Between Clusterings

Previous comparisons revealed that from a perceptual point of view, the synthesized friction sounds generated from recorded velocity profiles contained the same relevant information than the recorded ones, which seems to confirm that the velocity profile is the information which is perceptually relevant to recover a gesture through a friction sound. To completely validate this initial hypothesis, the perceptual and the *mathematical* dendrograms have been compared using the statistical method of cophenetic distances presented in the previous paragraph. The comparison reveals that for each corpus, except for the Pearson's correlation coefficient obtained for the recorded sounds of the experiment 1, the comparisons between the perceptual clusterings are not significantly different from the mathematical ones (see Table 4 for the Pearson's and Spearman's correlation coefficient). This result reinforces the importance of the velocity profile

in the perceptual process underlying the sound to shape association task, which was already suggested with the correlation between the perceptual dendrograms obtained from the listening tests.

Table 4. Pearson’s and Spearman’s correlations, respectively noted r and ρ , between the cophenetic distances obtained from the perceptual and the mathematical clusterings of the two experiments — Significant comparisons are bolded.

| | Experiment 1 | | Experiment 2 | |
|-------------|--------------|------------|--------------|------------|
| | r | ρ | r | ρ |
| Recorded | .58 | .93 | .98 | .93 |
| Synthesized | .89 | .93 | .94 | .93 |

3.4 Discussion

The clusterings reported here highlight the perceptual relevance of the velocity profile to evoke a graphical human movement. We firstly established an *objective* classification of shapes of two corpuses from the velocity recorded on a person drawing them. Shapes expected to have a close geometries are also close according to this metric like the ellipse and the loops for instance.

Our interest dealt with the auditory perception of gestural movements, and particularly to determine if the velocity of a gesture is perceptually relevant to characterize a gesture. We therefore compared the mathematical classification with perceptual ones thanks to the results of two listening tests, one for each corpus. The tests were composed of friction sounds recorded when the same person draws the shapes. Synthetic sounds generated only from the velocity were also used, which made it possible to investigate the perceptual relevance of the velocity. The variation of timbre involved in the recorded and in the synthesized sounds enabled the shape recognition. Finally, the comparisons between the perceptual and the mathematical classifications confirmed that the velocity profile of a gesture contains relevant information about the gesture underlying a friction sound. In particular that a sound can evoke a gesture. And even, to evoke a geometrical shape although the relation between a velocity profile and a shape is not bijective, i.e. one velocity profile can be the cause of the drawing movement of several geometrical shapes.

Henceforth we know that the velocity profile transmits sufficient information about the gesture, sufficient, to a certain extent, to discriminate different shapes from sounds. This implies that the kinematics of the gesture and the geometrical

characteristics of the drawn shape are correlated and gives an invariant information which enables subjects to extract a common representation of the gesture evoked by the sound, a *so-called* transformational invariant.

4 An Acoustical Characterization of a Human Gesture

When someone draws a shape on a paper, the final trace is static and all the dynamic information is lost *a priori*. The previous experiments pointed out that from an acoustical point of view, the velocity profile of a gesture was a relevant information to recognize a gesture from the friction sound produced when someone is drawing. It indeed enabled the discrimination of shapes when they were distinct. Conversely, when the shapes had close geometries, perceptual confusions between sounds (both for recorded and synthesized ones) appeared in particular for the ellipse and the loops. This result reveals that the gesture made to draw a shape is closely linked to its geometry and particularly to its kinematics.

4.1 The 1/3-Power Law

Many studies have already focused on this relation between the kinematics of a movement and the geometry of a trajectory. In particular, studies led by Paolo Viviani and his colleagues since the eighties highlighted that a biomechanics constraint implies the velocity of a gesture to depend on the curvature of the traveled trajectory [39]. Besides, they proposed a power law relation between the angular velocity of the pen and the curvature of the drawn trajectory [22, 40]. In terms of tangential velocity v_t and curvature C , it can be written:

$$v_t(s) = KC(s)^\beta \quad (3)$$

K is called the velocity gain factor and is almost constant during a movement. s is the curvilinear abscissa. The exponent β is close to 1/3 for adults' drawing [42] and the law has therefore been called the 1/3-power law. Possible description of this relation is that when we draw a shape, we accelerate in the flattest parts and we slow down in the most curved ones. This general principle constrains the production of biological movements but has also consequences in other sensorimotor modalities. Visual experiments revealed that a dot moving along a curved shape was perceived as the most constant when the relation between its velocity and the curvature of the traveled trajectory followed the 1/3-power law, even when the velocity variations exceeded 200% [43]. The relevance of this law has also been studied in the kinesthetic modality and revealed the same perceptual constraint [45].

This law can partly explain why the ellipse and the loops from the previous experiment were confounded. As their geometries were close, the velocities were also close, and as the produced friction sounds mainly depend on the velocity, they were not different enough to be distinguishable from an acoustical point of view. This law has therefore audible consequences, and it is legitimate to wonder if this biological invariant can be calibrated in the auditory modality.

4.2 Auditory Calibration

We adapted the protocol of visual calibration of the power law proposed in [43] to investigate the auditory perception of biological motion.

Subjects. Twenty participants took part in this experiment. Their average age was 29.42 years (SD=12.54).

Stimuli. While in the visual case studied in [43] the stimuli which were adjusted by subjects were moving visual dots, in the acoustical case, they were friction sounds. The previous synthesis model of friction sounds was used to generate a sound only from the kinematics (i.e. the velocity profile). These velocity profiles were computed by using the β -power law with a fixed mean velocity K . To avoid evoking specific known shapes, the curvature profiles were computed from pseudo-random shapes, see Figure 8 for an example.

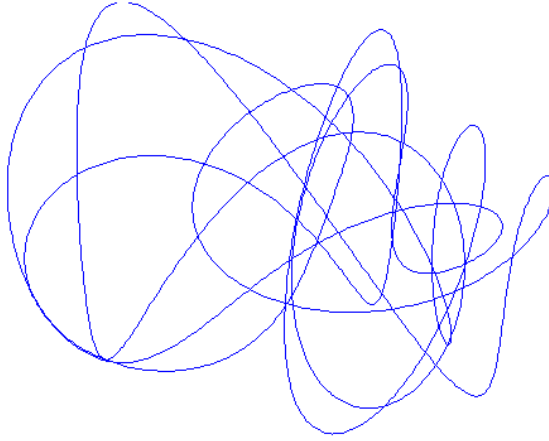


Fig. 8. An Example of Pseudo-Random Shape.

Task. Each subject effectuated 6 trials and a pseudo-random shape was generated for each trial. The subjects listened to the corresponding friction sound and were asked to adjust the sound until they perceived a sound which evoked the most fluid/natural movement. They could act on the timbre variations with two buttons which modified the β value. The subjects were unaware of the parameter on which they were acting. The initial value of β was randomized at each trial and the shape was not shown to the subjects to make them focus on the sound only.

Results and Short Discussion. The subjects adjusted the exponent value with a mean value of $\beta=0.36$ (SD=0.08), it is therefore not significantly different from the 1/3-power law ($p=0.10$). This indicates that to produce a sound which evoked a fluid and natural gesture, the velocity profiles from which a friction sound is generated should follow the 1/3-power law. This result reinforces the importance of kinematics in the perception and representation of gestures from a cognitive viewpoint. The sensorimotor relations between auditory and other modalities have more or less been investigated regarding the mental representation of gestures.

To conclude, one of our motivations was to control a friction sound synthesis model with intuitive controls. The research perspectives that we can expect made it possible to imagine a sound synthesis platform which enables to synthesize sounds from intuitive controls based on the 1/3-power law. The scope of such a platform is presented in figure 9. An interesting perspective of experiment 3 would be to ask the subject to adjust the sound (implicitly the exponent of the power law) evoking an aggressive and jerky gesture, or conversely the sound evoking a sweet caress, which are two gestures evoking different intentions. Intentions are closely linked to emotions and could be classified according to the classic valence-arousal scale [5]. Intuitively, the high parameter values of the power law would correspond to an aggressive and jerky gesture, i.e. to strong accelerations and strong decelerations at a high velocity. And conversely, for the caress, which involves *a priori* a slower and smoother gesture, the power law corresponding values would be small. Not especially null for the exponent, which corresponds to a uniform and constant movement and which has therefore no audible variations, but for values high enough to perceive a smooth and sweet gesture.

5 General Discussion

The starting point of this study was to investigate the relation between sounds and gestures, more specifically to understand whether a sound can evoke a gesture, and which characteristics could be conveyed by sound. In a general context, in the previous studies by Adrien Merer [27, ?], concerning the relation between an abstract sound and the evoked movements, the following questions were evoked: Why a sound evoke a specific movement? For instance, why does a sound evoke something oscillating or rotating? We therefore found it interesting to tackle this question in the case of human drawing movement regarding the sensorimotor considerations presented before.

By adopting the invariant's taxonomy, the results of the three previous experiments gave important cues about one *transformational invariant* which characterizes a gesture: its kinematic. Henceforth we know that the velocity transmits relevant information about the gesture, which moreover can be associated to a geometric shape to a certain extent. At last, the third experiment brings to light the relevance of a biological relation between the velocity and the curvature from the auditory point of view, which has never been investigated before. This last

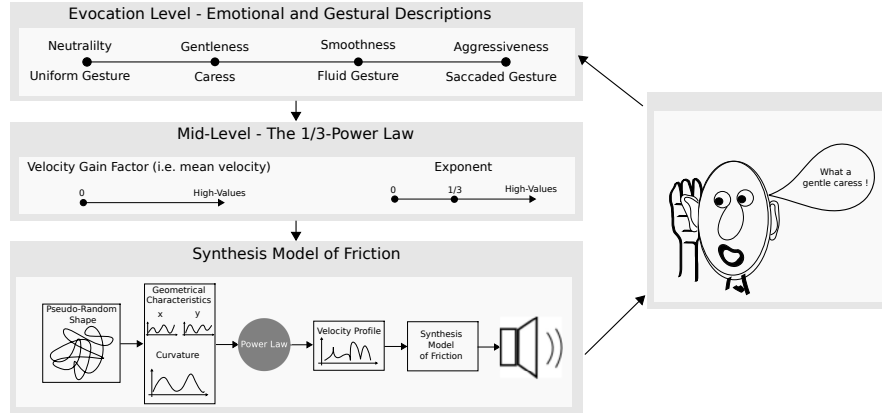


Fig. 9. Architecture of the friction sound synthesis platform with intuitive controls of the evocation. The controls discussed in the conclusion such as, aggressiveness and gentleness, have been proposed. A formal calibration of the velocity gain factor and the exponent with respect to the gestural and emotional descriptors will be conducted in future works.

experiment also showed the interesting result that, to evoke a fluid and natural gesture through a friction sound, the velocity profile should follow the $1/3$ -power law, which means that friction sounds could directly informed about the naturalness and the fluidity of a gesture. This point has to be discussed because it opens the possibility to recognize qualities of a human gesture through sounds, which provides new perspectives in the understanding of auditory perception and its relation with other modalities.

Going back to the task of the experiment 3, it is interesting to ask ourself what it involves to ask someone to adjust a friction sound to obtain the most fluid and natural sound according to a human gesture? In a representationalist view, it firstly means that our perceptual system has to extract an information from the timbre variations – mainly the brightness in the case of friction sounds – which is then abstracted and *internalized* as the velocity. And secondly, to compare it with an internal representation of fluidity to decide if this velocity corresponds to a fluid gesture, and eventually to change a parameter of the sound and to start the process again. This view is not trivial and supposes that we have internal representations of gestures – and moreover fluid gesture – which can be compared with the one computed from an incoming auditory stimulus. In the case of experiment 1 and 2, we are even able to associate this dynamic representation of the gesture to a geometrical one, the static visual shape. The problem of representations in perception has been widely discussed in the visual modality to understand how a physical system, such as the brain, can makes the feeling of seeing, which is not *a priori* a physical state. The enactive approach of Varela presented in the introduction placed the action in the center of the perceptual processes. The sensorimotor theory of perception of Kevin O'Regan

[31] proposed an interesting approach of this assumption. It is argued that *seeing is not making an internal representation of the outside world from the visual input*, in other words to make a mirror in the brain of the world from a visual stimulus. It is proposed that *seeing is knowing about things to do* with the outside world, i.e. knowing the actions you can make with the outside world and their consequences on your sensory input. For instance, seeing a wine glass is not having a picture of it in the head, but it is projecting what you can do with it, filling it or drinking it for example. Based on behavioral and neurophysiological observations, Marc Jeannerod and Alain Berthoz respectively introduced the notion of *Simulated Action – the perception is a simulated action*, [16, 4, 17]. It sums up the idea that perceiving, whatever the modality, is extracting the relevant information to interact with the objects of perception, e.g. grasping the glass to drink it. It is therefore making hypothesis about what and how I can interact with it according to my motor competences. For example, when we see a tea cup on a table, we are able to say which hand is the most adapted to grasp it according to the position of the handle, the right hand if the handle turn to right and conversely. Simulated action seems to involve the same processes as the one proposed in the sensorimotor theory of O'Regan.

If we apply the sensorimotor approach to the auditory perception in the case of the sounds produced by drawing movements, the same distinction as in the visual case can be made. Listening to the friction sounds produced by someone who is drawing is not making an internal representation of the sound produced by the gesture, but it is imagining executing the corresponding gesture to which the acoustical consequences are the perceived sound with the same timbre variations. All the action planning is involved in this task, from the intention to the proactive simulation of the execution of the movement. Finally, according to the sensorimotor paradigm, to perceive a friction sound it is almost already doing the gesture which has produced the sound. This distinction is interesting because it gives a relevant approach to make hypothesis about the invariant features which enable the auditory recognition of acoustical events. Regarding the previous definition of a simulated action, the third experiment reinforces this notion from the auditory point of view. The subjects have been able to take the place of the writer and to adjust the sound by mentally simulating the gesture they would have executed and to compare it with the internal reference of fluid gestures. Understanding the behavior of human beings involves understanding their actions, intentions, and even emotions, to react and to behave appropriately. The main difference between humans and animals is definitely the existence of a highly developed language, which enables sharing of actions, intentions and emotions through a common channel of communication among individuals. An hypothesis widely accepted now is that our verbal language derived from a gestural one, in which actual words were screams and contortions [33]. Nowadays, the ability of humans to speak articulately enables to use quasi exclusively the vocal channel rather than the gestural one. But we have not completely abandoned the gestures, and it is commonly observed that when we speak, we make a lot of gestures with a very important signification according to the context to supply

our speech [30]. Another observation about gestural language and more generally about corporeal language, is related to the postural communication in animals. Darwin studied the dog postures and their significations, and remarked that posture can either evoke hostile intents or humble and affectionate mood [9]. Hence gestures and more generally corporeal articulations have a lot of perceptual significations, in line with the third experiment previously presented from the auditory modality which will be discussed.

In experiment 3, we proposed a perspective of intuitive controls based on emotional and gestural descriptions. The close relation between intention, emotion and gesture was addressed in a musical context, this question has been addressed by Marc Leman [23], a proposition to explain interpretations and emotions involved by musical and aesthetic experiences by considering the *corporeal engagement*, and the *corporeal resonances*, which are non linguistic descriptions of music and involved emotions. Such corporeal engagement has been suggested in the case of the emotions we feel when we see a painting. It could be due to the fact that we try to imagine the gestures, and therefore the underlying intentions of the artist. This question has been already discussed in [26] regarding functional imagery studies which revealed that our motor areas involved in drawing were also activated when we see passively a letter we are able to trace [24, 25]. These results suggest that the visual perception of a stimulus involved all the processes implied in the motor planning, and could be the basis of emotions engaged when we perceive an artistic performance. Jeannerod suggests it in the case of the perception of a dance performance [17]. He proposed that we may perceptually take the place of the dancer in order to feel his sensorial state, so the emotions we can feel from a dance performance could be explained by such a perceptual process coupling perception to simulated action. By analogies with such processes involved in vision and according to the results of the experiment 3, we could imagine that such simulated actions should also be involved in the case of the perception of a musical piece, but more experiments, either behavioral or from functional imagery should be done to confirm such a strong hypothesis.

Finally, the enactive theory and the notion *simulated action* are a well adapted framework for studying the perception of auditory objects which involved a human action like the friction sounds produced when someone is drawing. Moreover, the invariant taxonomy seems also well adapted to this. But to extend these approaches to environmental sounds, which do not necessarily involved an embodied action, we have to define a general concept of *action* and to take into account our interaction with the surrounding world with an holistic point of view. For instance, the sound of a river flowing does not involve a human action, in the sense of producing a motor act such as drawing or making a movement. And finally, what is simulated when we listen such sound? It is maybe more generally linked to experience rather than a simulated motor action. And maybe it would be interesting to define a notion which could be named a simulated situation, which englobes embodied active actions but also contains more general experiences. This point will not be discussed here but have to be clarified to establish a general ontology of sound perception based on the sensorimotor

and phenomenal contingencies. It would also be interesting to discuss and to contextualize the invariants taxonomy regarding this more general framework.

6 Acknowledgments

This work was funded by the French National Research Agency (ANR) under the MetaSon: Métaphores Sonores (Sound Metaphors) project (ANR-10-CORD-0003) in the CONTINT 2010 framework. The authors are grateful to Charles Gondre for his precious help for the development of the listening test interfaces.

References

1. Aramaki, M., Gondre, C., Kronland-Martinet, R., Voinier, T., & Ystad, S.: Thinking the sounds: an intuitive control of an impact sound synthesizer. *Proceedings of ICAD 09 – 15th International Conference on Auditory Display* (2009)
2. Aramaki, M., Besson M., Kronland-Martinet R., & Ystad, S.: Controlling the Perceived Material in an Impact Sound Synthesizer. *IEEE Transactions On Speech and Audio Processing*, 19(2): 301–314 (2011)
3. Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H. J. & Altenmüller, E.: Shared network for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage*, 30(3), 917–926 (2006)
4. Berthoz, A.: *Le sens du mouvement*, Odile Jacob (1997)
5. Bigand, E., Vieillard, S., Madurell, F. and Marozeau, J., & Dacquet, A.: Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, 19(8) 1113–1139 (2005)
6. Carello, C., Anderson, K. L. & Kunkler-Peck, A. J.: Perception of object length by sound. *Psychological Science*, 9(3), 211–214 (2008)
7. Conan, S., Aramaki, M., Kronland-Martinet, R., Thoret, E., & Ystad, S.: Perceptual Differences Between Sounds Produced by Different Continuous Interactions. *Acoustics 2012*, Nantes (2012)
8. Conan, S., Aramaki, M., Kronland-Martinet, R., Thoret, E., & Ystad, S.: Intuitive Control of Rolling Sound Synthesis. In *Aramaki, M., Barthelet, M., Kronland-Martinet, R., & Ystad, S. (Eds.): CMMR 2013*, volume 7900 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (2013)
9. Darwin, C.: The expression of the emotions in man and animals, *The American Journal of the Medical Sciences*, 232(4), 477 (1956)
10. Frémiot, M., Mandelbrojt, J., Formosa, M., Delalande, G., Pedler, E., Malbosc P., & Gobin, P.: Les Unités Sémiotiques Temporelles: éléments nouveaux d’analyse musicale. Diffusion ESKA. MIM Laboratoire Musique et Informatique de Marseille (MIM), documents musurgia édition. 13 (1996)
11. Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G.: Action recognition in the premotor cortex, *Brain*, 119(2), 593–609, (1996)
12. Gaver, W. W.: What in the world do we hear?: an ecological approach to auditory event perception. *Ecological psychology*, 5(1) 1–29 (1993)
13. Gaver, W. W.: How do we hear in the world? Explorations in ecological acoustics. *Journal of Ecological Psychology*, 5(4), 285–313 (1993)

14. Griffiths, T. D., & Warren, J. D.: What is an auditory object?. *Nature Reviews Neuroscience*, 5(11), 887–892 (2004)
15. Gibson, J. J.: *The Senses Considered as Perceptual Systems*. Boston: Houghton-Mifflin (1966)
16. Jeannerod, M.: The representing brain: Neural correlates of motor intention and imagery, *Behavioral and Brain sciences*, 17(2), 187–202 (1994)
17. Jeannerod, M.: *La fabrique des idées*, Odile Jacob (2011)
18. Klatzky, R. L., Pai, D. K., & Krotkov, E.P.: Perception of material from contact sounds. *Presence: Teleoperators & Virtual Environments*, 9(4), 399–410 (2000)
19. Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V., & Rizzolatti, G.: Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297(5582), 846–848 (2002)
20. Keogh, E., & Ratanamahatana, C.A.: Exact indexing of dynamic time warping, *Knowledge and information systems*, 7(3), 358–386 (2005)
21. Kronland-Martinet, R., & Voinier, T.: Real-time perceptual simulation of moving sources: application to the Leslie cabinet and 3D sound immersion. *EURASIP Journal on Audio, Speech, and Music Processing* (2008)
22. Lacquaniti, F., Terzuolo, C. A., & Viviani, P.: The law relating kinematic and figural aspects of drawing movements, *Acta Psychologica*, 54, 115–130 (1983)
23. Leman, M.: *Embodied music cognition and mediation technology*, Mit Press (2007)
24. Longcamp, M., Anton, J.L., Roth, M., & Velay, J. L.: Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage*, 19(4), 1492–1500 (2003)
25. Longcamp, M., Tanskanen, T., & Hari, R.: The imprint of action: Motor cortex involvement in visual perception of handwritten letters. *Neuroimage*, 33(2), 681–688 (2006)
26. Longcamp, M., & Velay, J.L.: Connaissances motrices, perception visuelle et jugements esthétiques: lexemple des formes graphiques, in *Dans l'atelier de l'art Expériences Cognitives* by Mario Borillo. Champ Vallon, Seyssel, France, 2010, 284, MIT Press (2010)
27. Merer, A., Ystad, S., Kronland-Martinet, R. & Aramaki, M.: Semiotics of sounds evoking motions: Categorization and acoustic features. In *Kronland-Martinet, R., Ystad, S., & Jensen, K., (Eds.): CMMR 2007. Sense of Sounds*, 139158. Springer, LNCS vol. 4969 (2008)
28. Merer, A., Aramaki, M., Ystad, S., & Kronland-Martinet, R.: Perceptual characterization of motion evoked by sounds for synthesis control purposes. *ACM Transaction on Applied Perception*, Col.10(1), February 2013 (2013)
29. McAdams, S.: Recognition of sound sources and events. In *McAdams, S., & Bigand, E. (Eds.): Thinking in Sound: The cognitive psychology of human audition*, 146198 (1993)
30. McNeil, D.: So you think gestures are nonverbal?. *Psychological review*, 92(3), 350–371 (1985)
31. O'Regan, J.K., & Noë, A.: A sensorimotor account of vision and visual consciousness, *Behavioral and brain sciences*, 24(5), 939–972 (2001)
32. Rizzolatti, G., Fadiga, L., & Gallese, V., & Fogassi, L.: Premotor cortex and the recognition of motor actions, *Cognitive brain research*, 3(2), 131–141 (1996)
33. Rizzolatti, G., & Sinigaglia, C.: *Les neurones miroirs*, Odile Jacob (2008)
34. Rosenblum, L.D., Carello, C., & Pastore, R. E.: Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception*, 16, 175–186 (1987)
35. Sokal, R. R., & Rohlf, F.J.: The comparison of dendrograms by objective methods. *Taxon*, 11(2), 33–40 (1962)

36. Van Den Doel, K., Kry, P. G. & Pai, D. K.: FoleyAutomatic : physically-based sound effects for interactive simulation and animation. *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 537–544, ACM (2001)
37. Varela, F. J.: Invitation aux sciences cognitives, Eds. Le Seuil (1988)
38. Varela, F. J., Thompson, E. T., Evan, T. & Rosch, E.: The embodied mind: Cognitive science and human experience, MIT press (1992)
39. Viviani, P., & Terzuolo, C.: Trajectory determines movement dynamics. *Neuroscience*, 7, 2, 431–437 (1982)
40. Viviani, P., & McCollum, G.: The relation between linear extent and velocity in drawing movements. *Neuroscience*, 10(1), 211–218 (1983)
41. Viviani, P., & Stucchi, N.: The effect of movement velocity on form perception: Geometric illusions in dynamic displays. *Attention, Perception, & Psychophysics*, 46(3), 266–274 (1989)
42. Viviani, P., & Schneider, R.: A developmental study of the relationship between geometry and kinematics in drawing movements. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 198–218 (1991)
43. Viviani, P., & Stucchi, N.: Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 603–623 (1992)
44. Viviani P., & Flash T.: Minimum-jerk, two-thirds power law and isochrony: Converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 32–53 (1995)
45. Viviani, P., Baud-Bovy, G., & Redolfi, M.: Perceiving and tracking kinesthetic stimuli: Further evidence of motor-perceptual interactions, *Journal of Experimental Psychology: Human Perception and Performance*, 23(4), 1232–1252 (1997)
46. Warren, W. H., & Verbrugge, R. R.: Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), 704 (1984)
47. Wildes, R. P., & Richards, W.A.: Recovering material properties from sound. In W. A. Richards (Ed.): *Natural computation*, 356–363 (1988)